

Book Review

**Christine Horne: The Rewards of Punishment.
A Relational Theory of Norm Enforcement**

Stanford 2009: Stanford University Press.
ISBN: 9870804760225

by Sonja Vogt

In her book, Horne develops a theory of norm enforcement. Norms are fundamental to explain social order. Norm enforcement has been studied in many ways in the social sciences. However, Horne adds a truly unique approach. She links norm enforcement with social interdependence by asking how norm enforcement relates to the social relations people have with each other. She defines interdependence as the extent to which people value their relations and the goods they can receive through relations with others. The book claims that the more interdependent people are, the more likely they sanction norm violations. This is similar to Coleman's claim that closed networks are essential for the realization of norms (1990, Chapter 11). Horne discusses norm and meta-norm enforcement, with multiple real world examples, and examines enforcement with laboratory experiments.

Before Horne introduces her own theoretical approach to explain norm enforcement she discusses various theoretical approaches, namely the consequentialist approach, the shared-meanings approach, and the typicality approach. The consequentialist approach claims that the smaller the costs and the larger the benefits of norm compliance, the more likely people enforce norms. The shared-meanings account merely claims that people agree on a particular behavior that they then all follow. The last approach, the typicality approach, is based on the assumption that individuals imitate the behavior of others due to normative pressure created by the desire to follow the majority. Horne nicely shows the limitations of each of the three approaches with respect to explaining norm and meta-norm enforcement. In order to give a more satisfactory account, Horne develops a relational theory of norm enforcement. The main point of the relational approach is that people not only sanction in response to bad behavior, they also sanction because they care what people think of them. People sanction others because they want to look good to others and not just because of future consequences of others' deviant behavior. The book covers some very important questions related to norm enforcement, and it gives interesting theoretical answers by linking norm enforcement with the relationships among individuals who either do or do not enforce social norms. What makes her approach so

interesting is that she does not study norm enforcement in terms of promised rewards and threatened punishment. Rather, she studies norm enforcement in terms of actual rewards and real punishment. In spite of this, however, a more thorough discussion of how her contribution relates to the existing literature on punishment and pro-social behavior (e.g. Fehr and Gächter 2002) would have made her analysis more accessible.

Horne tests her theoretical claims with experimental data. To test her theory she develops the Norms Game, the Metanorms Game, and the Expectations Game. Most of her hypotheses are seemingly confirmed. However, the empirical findings are often difficult to interpret. The *Norms Game* tests whether “people impose harsher punishments against behavior that produces greater harm and whether this tendency is enhanced when group members are interdependent” (28). In the Norms Game subjects interacted with each other in groups of four for 48 rounds. They made decisions about donating to a group project. Each round contained four stages. In the first stage every subject received an income of 30 points. In the second stage, one randomly chosen subject decided whether or not to donate 20 points to the group and as a consequence everybody, including the donating subject, received a certain amount of points. In the third stage subjects had the opportunity to give points to each other. In the final stage, each subject received her points in her account, and the game proceeded to the next round. The experiment contained two main treatment variations in a 2x2 design, namely the consequences of donating points in the second stage were either low or high, and the interdependence among subjects in the third stage was either low or high. In the consequences dimension, subjects either received 2 or 6 points given one member of the group donated 20 points in the second stage. Thus, the treatment difference was the difference in the points subjects received after a donation happened. In the interdependence dimension, Horne manipulated the way the points were exchanged between subjects in stage three. In the high dependency case, the experimenter doubled the points exchanged among subjects. If player W gave away 10 points to player Y, player Y actually received 20 points. In the low interdependence case, points were exchanged, but the experimenter did not double them.

Interdependence was manipulated in a way such that the low interdependence treatment resulted in no efficiency gains or losses. The high interdependence treatment, in contrast, produced efficiency gains because the exchanged points were doubled and therefore new money entered the experiment; it was efficient to exchange. It is therefore not clear whether it was the interdependence among subjects or the differences in efficiency that caused peoples’ behavior to change. It has been shown in several studies that some people are motivated by efficiency gains (e.g. Charness and Rabin 2002). Efficiency motives and relational motives can both affect subject behavior in Horne’s experiment based on the Norm Game. If the motivation of the behavior in the second stage is not clear, one has to be careful with the interpretation of the results. Even though the analysis shows that subjects exchanged more in the third stage with those

subjects who donated in the second stage, the effects of interdependence cannot be isolated if they are confounded with efficiency.

Efficiency concerns also complicate the interpretation of effects associated with the consequences dimension. In the low consequence treatment, subjects suffered a small efficiency loss when donating money (20 points donated, but only 8 points received by the group in total). In the high consequence treatment, subjects had a small efficiency gain (20 points donated, and 24 points received by the group in total). In other words, donating money in the low consequence treatment burned money, while donating money in the high consequence treatment created money. In the low consequence treatment, donating points was framed as a donation to the group, but in reality it resulted in both a cost to the donor and an efficiency loss for the group. Thus, it is not clear whether donating was really prosocial. In the high consequence treatment, donating was costly for the individual, but it produced an efficiency gain. Again, the design does not allow to disentangle efficiency motives from prosocial motives.

A more general point regarding the design of the Norms Game is the assumption that behavior in the third stage of the game was related to behavior in the second stage. It would be important to provide clear evidence for this with a manipulation check on showing the link between donation and exchange.

Horne does a similar study focusing on meta-norm enforcement rather than norm-enforcement. Meta-norms are defined as second-order norms that regulate the sanctioning process of norm violations. In line with the norm enforcement study beforehand, Horne predicts that meta-norm enforcement is more likely the more interdependent the people are. Horne tests her prediction by developing a *Metanorms Game*. Subjects were in groups of five. One of them was randomly assigned to a situation where a thief stole something. The victim could then decide to punish the thief. Punishing was costly for the victim, but at the same time it benefited the others in the group. After the victim decided whether or not to punish, all subjects in the group could exchange points as described in the Norms Game above. In the high interdependence treatment the experimenter tripled exchanged points and in the low interdependence treatment points were kept constant. In addition to interdependence, Horne manipulated the presence or absence of indirect sanctioning benefits. The points subjects received if a thief was punished were either used in the game for exchange (indirect benefit), or they were locked up and paid at the end of the game (direct benefit). Horne's predictions were confirmed, but again, as in the Norms Game, interpreting the results is not always straightforward because interdependence and efficiency are confounded.

The most innovative chapter of Horne's book is about expectations and meta-norm enforcement. In a nutshell, Horne claims that people often sanction behavior, even when the behavior does not harm anybody, simply because the behavior is non-conformist. A classical example she gives is fashion. She explains punishing atypical behavior by arguing that people believe the punishment of deviations, even if deviations are costless, will be seen favorably by others. This is a very interesting claim and Horne tests it with the *Expectation Game* she

developed. The Expectation Game is a variant of the classic Asch (1951) experiment. In Horne's version, the actors chose between two letters, X and W. Each subject was in a group with seven simulated actors. Real subjects were not told that the other subjects were simulated. After six of the simulated actors made a choice that was actually programmed (either all chose X or W), the real subject had to make a choice. The simulated actor who then followed the real subject made a choice that differed from the formerly simulated choices. Then the real subject had the opportunity to punish the deviant simulated subject by giving more points to those who conformed with the majority relative to those who did not. Horne hypothesized that real subjects who conformed to the majority decision would punish those sequent choices that deviated. Interestingly, Horne did not find her predicted effect initially. Only when she gave some kind of meaning to the behavioral choices did she get the predicted effects. In this case, the real subject was told that choosing X or W was related to the number of friends the simulated subjects had. This is an interesting, though unexplained result, and it would be useful to study the reason for the framing effect.

In the final chapters of the book, Horne argues that her theory applies to a range of norm enforcing behaviors outside the laboratory. It is important to validate laboratory findings with empirical tests outside the experimental laboratory, and norm enforcement is a potentially rich topic for connecting experimental work with survey data. Horne discusses examples associated with human rights, deviant behavior, and norm enforcement in neighborhoods. Altogether, *The Rewards of Punishment* makes an interesting contribution to the study of norm enforcement. Horne introduces refreshing ideas by studying the effects of social relations on norm enforcement and meta-norm enforcement.

References

- Asch, S. (1951), "Effects of Group Pressure upon the Modification and Distortion of Judgements", in: H. Guetzkow (ed.), *Groups, Leadership, and Men*, Pittsburgh: Carnegie Press, 177–190.
- Charness, G. and M. Rabin (2002), "Understanding Social Preferences with Simple Tests", *The Quarterly Journal of Economics* 117(3), 817–869.
- Coleman, J. S. (1990), *Foundations of Social Theory*, Cambridge/MA: Belknap Press of Harvard University Press.
- Fehr, E. and S. Gächter (2002), "Altruistic Punishment in Humans", *Nature* 415, 137–140.